라그랑주 승수법

I. Intro

라그랑주 승수법(Lagrangian Multiplier Method)은 제약 조건이 있는 최적화 문제를 풀기 위해 고안되었다.

라그랑주 승수법을 알아보기 전에 먼저 Duality에 대한 내용을 살펴보자.

II. Duality

최적화를 요하는 문제에서 duality란, 어떠한 최적화 문제를 primal problem(원초문제)과 dual problem(쌍대문제)의 두 가지 관점에서 볼 수 있다는 뜻이다. 예를 들어 살펴보자.

EXAMPLE

$$\min_x \quad c^T x$$
$$subject\ to$$
$$Ax = b \quad and \quad Gx \leq h$$

위처럼 주어진 선형 조건을 만족하면서 동시에 선형함수인 $c^T x$를 최소화하는 문제가 있다.

여기서 c는 n차원, b는 m차원, h는 r차원 벡터, A는 $m \times n$, G는 $r \times n$ 행렬이다.

이를 스칼라 형태의 식으로 살펴보면 등식 형태의 제약식이 m개, 부등식 형태의 제약식이 r개 있고, 이를 만족하며 목적함수를 최소화 하는 n개의 미지수를 찾는 문제라고 볼 수 있다.

두 가지 제약식의 양변에 $u,\ v$라는 벡터를 곱해보자.
이 때 벡터 v의 요소가 모두 0 이상의 값을 갖게 해 부등식의 방향이 바뀌지 않도록 한다.

$$u^T Ax = u^T b$$
$$v^T Gx \leq v^T h$$

위 두 식을 각 변끼리 더하고 정리하면 다음과 같다.

$$\begin{gathered} u^TAx + v^TGx \leq u^Tb + v^Th \\ (A^Tu + G^Tv)^Tx \leq u^Tb + v^Th \\ (-A^Tu - G^Tb)^T \geq -u^Tb - v^Th \\ \therefore \quad c^Tx \geq -u^Tb - v^Th \end{gathered}$$

위에서 볼 수 있듯이 $-A^Tu - G^Tb = c$가 되도록 하면 primal problem의 목적함수 c^Tx의 하한은 $-u^Tb - v^Th$가 된다.

따라서 위의 primal problem과 아래의 dual problem은 쌍을 이루게 된다.

$$\begin{aligned} \max_{u,v} \quad & -u^Tb - v^Th \\ subject\ to \quad & -A^Tu - G^Tb = c,\ and\ v \geq 0 \end{aligned}$$

III. 라그랑주 승수법

1. 기하학적인 이해

이제 라그랑주 승수법을 알아보도록 하자.

라그랑주 승수법의 기본 가정은 **"제약조건 g를 만족하는 f의 최소값 또는 최대값은 f와 g가 접하는 지점에 존재할 수도 있다."**는 것이다.

라그랑주 승수법에서는 f와 g가 접하는 지점을 찾기 위해 gradient vector를 이용한다고 한다. Gradient vector는 각 미지수로 편미분한 값을 원소로 갖는 벡터이므로 $\nabla f = (\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y})$의 꼴로 나타낼 수 있다.

그리고 접선 벡터와 gradient vector의 dot product는 0, 즉 수직 관계이므로 두 함수 f와 g가 접한다는 것은 두 함수의 gradient vector가 서로 상수배의 관계임을 의미한다. 이를 다음과 같이 나타낼 수 있다.

$$\nabla f = \lambda \nabla g \quad (1)$$

라그랑주 승수법에서는 라그랑지안 함수 L을 다음과 같이 정의한다.

$$L(x, y, \lambda) = f(x, y) + \lambda(g(x, y) - c)$$

함수 L의 gradient vector가 0이 되는 지점을 찾으면 (1)의 꼴이 되기 때문에 두 함수 g와 f가 접하는 점을 찾을 수 있게 된다.

따라서 함수 L을 x와 y에 대해 편미분하고 이에 더해서 $g(x, y) = c$를 만족하는 조건을 이용하면 위 문제의 solution을 구할 수 있다. 그리고 이렇게 구해진 x와 y는 제약조건 g를 만족하는 함수 f의 최적점이 될 가능성이 있게 된다.

2. 전미분을 이용한 해석

어떤 함수 $f(x, y, z)$의 최솟값 혹은 최댓값은 극점에 존재할 수 있고, 다변수 함수의 극점은 전미분 했을 때 그 값이 0인 지점에 존재한다.

다변수 함수의 전미분은 아래와 같이 정의된다.

$$df = \frac{\partial f}{\partial x}dx + \frac{\partial f}{\partial y}dy + \frac{\partial f}{\partial z}dz$$

변수 dx, dy, dz가 각각 독립이라면 위 식을 0으로 만드는 조건은

$$\frac{\partial f}{\partial x} = \frac{\partial f}{\partial y} = \frac{\partial f}{\partial z} = 0$$

과 같다.

위 함수 f의 최적해를 찾기 위한 제약조건 $g(x, y, z) = 0$에 대해서 전미분 하면 아래와 같다.

$$dg = \frac{\partial g}{\partial x}dx + \frac{\partial g}{\partial y}dy + \frac{\partial g}{\partial z}dz = 0$$

그리고 위 식을 다시 dz에 대해 정리하면

$$dz = -\frac{\frac{\partial g}{\partial x}dx + \frac{\partial g}{\partial y}dy}{\frac{\partial g}{\partial z}}$$

가 되고 이를 위 식에 다시 대입해 정리하면

$$df = \frac{\partial f}{\partial x}dx + \frac{\partial f}{\partial y}dy - \frac{\partial f}{\partial z}\frac{\frac{\partial g}{\partial x}dx + \frac{\partial g}{\partial y}dt}{\frac{\partial g}{\partial z}} = 0$$

을 해결함으로써 f의 전미분이 0이 되는 지점을 찾을 수 있다.

어떤 상수 λ를 $\lambda = \frac{\frac{\partial f}{\partial z}}{\frac{\partial g}{\partial z}}$로 두고 위의 식을 정리하면

$$(\frac{\partial f}{\partial x}dx + \frac{\partial f}{\partial y}dy) - \lambda(\frac{\partial g}{\partial x}dx + \frac{\partial g}{\partial y}dy) = 0$$

이 되고 dx와 dy가 독립을 만족하면서 위 식이 해결되려면

$$\frac{\partial f}{\partial x} = \lambda\frac{\partial g}{\partial x}$$
$$\frac{\partial f}{\partial y} = \lambda\frac{\partial g}{\partial y}$$

와 같은 식이 성립해야만 한다.

이는 위의 기하학적인 해석에서 함수 f의 gradient vector와 제약 조건 g의 gradient vector가 서로 스칼라 곱의 관계라는 점과 동일하다.

IV. 쌍대문제에 대한 라그랑지안적 접근

이제 라그랑주 승수법에 대한 이해를 바탕으로 처음에 본 Duality로 돌아가보자.

$$\min_x \quad c^T x$$
$$subject\ to$$
$$Ax = b \quad and \quad Gx \leq h$$

위와 같은 primal problem에 대해서 라그랑주 승수 벡터 u와 v를 도입해서 라그랑주 함수 L을 만들면 아래와 같은 형태가 나온다.

$$L(x, u, v) = c^T x + u^T(Ax - b) + v^T(Gx - h) \leq c^T x$$

그런데 이 때 두 번째 항은 무조건 0이고 ($Ax = b$), 세 번째 항은 반드시 0보다 작다. 따라서 L은 우리의 목적함수 $c^T x$보다 반드시 작거나 같게 되는 것이다.

제약 조건이 있는 상태보단 없는 상태일 때 더욱 작은 최적값을 얻을 수 있을 것은 자명하다(최소화 문제일 때). 이를 부등식으로 표현하면 아래와 같다.

$$f^* \geq \min_{x \in C} L(x, u, v) \geq \min_x L(x, u, v)$$

f^*은 우리가 찾으려는 최적값, C는 primal problem의 제약식을 만족하는 x의 집합이다.

이 때 가장 마지막 항($g(u, v)$로 표현하겠다.)은 그 정의 상 라그랑주 함수 L의 최소값이고, 함수 L은 당연히 우리가 알고 싶은 미지수로 편미분한 결과가 0이 되는 지점에서 최소값을 가질 것이다. 위에서 우리가 알고 싶은 미지수는 x이므로 그것에 대해 편미분 하면

$$
\begin{gathered}
\frac{\partial L}{\partial x} = c^T + u^T A + v^T G = 0 \\
\therefore \quad c = -A^T u - G^T v
\end{gathered}
$$

위 식을 L에 대입해서 풀면 $g(u, v)$를 구할 수 있다.

$$
\begin{gathered}
L(x, u, v) = c^T x + u^T (Ax - b) + v^T (Gx - h) \\
\Rightarrow (-A^T u - G^T v)^T x + u^T (Ax - b) + v^T (Gx - h) \\
= -u^T Ax - v^T Gx + u^T Ax - u^T b + v^T Gx - v^h \\
= -u^T b - v^T h \\
= g(u, v)
\end{gathered}
$$

우리는 이미 $f^* \geq g(u, v)$라는 관계를 확인했기 때문에, primal problem의 제약조건을 만족하는 최소값(최소화 문제이므로 최적값 = 최소값)을 찾는 것은 곧 $g(u, v)$를 최대화 하는 문제와 동일하다는 사실을 확인할 수 있다.

단, 라그랑주 승수법을 적용하는 과정에서 가정한 두 가지 조건인 $-A^T u - G^T v = c, \ v \geq 0$을 만족해야 한다.

따라서 우리는 primal problem과 쌍을 이루는 dual problem을 라그랑주 승수법을 이용해 정의할 수 있다.

$$
\begin{aligned}
\max_{u,v} \quad & -b^T u - h^T v \\
subject\ to \quad & -A^T u - G^T v = c \quad and \quad v \geq 0
\end{aligned}
$$

원초 문제를 쌍대 문제로 바꾸는 과정에서 라그랑지안적 접근의 장점은 예시로 들었던 linear programming이 아닌 임의의 최적화 문제에서도 적용이 가능하다는 점이다. 최적화 문제를 일반화 해보도록 하겠다.

$$
\begin{gathered}
\min_x \quad f(x) \\
subject\ to \\
h_i(x) \leq 0, \quad i = 1, ..., m \quad and \\
l_j(x) = 0, \quad , j = 1, ..., r
\end{gathered}
$$

이에 대해서 라그랑주 함수 L과 라그랑지 듀얼 함수 $g(u, v)$를 정의해보자. 앞에서 보았듯이 라그랑주 승수 u_i는 반드시 0보다 커야 한다. (부등식 꼴의 제약조건에서 부등호의 방향을 유지하기 위함)

$$L(x,u,v) = f(x) + \sum_{i=1}^{m} u_i h_i(x) + \sum_{j=1}^{r} v_i l_i(x)$$
$$g(u,v) = \min_x L(x,u,v)$$

다시 한 번 $f^* \geq g(u,v)$이므로 dual problem을 아래와 같이 쓸 수 있다.

$$\max_{u,v} \quad g(u,v)$$
$$subject\ to \quad u \geq 0$$

따라서 주어진 조건을 만족하는 최적의 벡터 x를 구하는 문제가 벡터 u, v를 구하는 쌍대 문제로 바뀌었다. 또한, 최소 문제가 최대 문제로 바뀐 점도 알 수 있다.

V. KKT Condition

Karush-Kuhn-Tucker 조건은 primal problem과 dual problem의 관계로부터 도출된 조건으로, 최적화 이론과 그것의 구현에서 핵심적인 역할을 한다고 한다. 먼저 그 정의부터 살펴보도록 하자.

$$Given\ general\ problem,$$
$$\min_x \quad f(x)$$
$$subject\ to$$
$$h_i(x) \leq 0, i = 1, ..., m$$
$$and \quad l_j(x) = 0, j = 1, ..., r$$

$$The\ KKT\ conditions\ are:$$

- $0 \in \partial(f(x) + \sum_{i=1}^{m} u_i h_i(x) + \sum_{j=1}^{r} v_j l_j(x)) \quad (stationary)$
- $u_i \cdot h_i(x) = 0 \quad for\ all\ i \quad (complementary\ slackness)$
- $h_i(x) \leq 0,\ l_j(x) = 0 \quad for\ all\ i,\ j \quad (primal\ feasibility)$
- $u_i \geq 0 \quad for\ all\ i \quad (dual\ feasibility)$

첫 번째와 세 번째 항목은 등식의 조건만 존재하는 최적화 문제의 경우와 동일하다.

그러나 두 번째와 네 번째 항목은 부등식의 조건이 존재하는 경우에만 나타나는 추가적인 조건이다.

쉽게 풀어 쓰면 부등식 제약조건에 관련된 라그랑주 승수는 음수가 되면 안 되고(네 번째 항목),
부등식 제약조건에 대한 라그랑주 승수와 해당 부등식의 곱은 반드시 0이 되어야 한다는 의미이다.

KKT 조건을 만족하는 x^*, u^*, v^*가 있다면, 이 때 x^*는 primal problem의 최적해가 되고, u^*, v^*는 dual problem의 최적해가 된다. 따라서 부등식이 존재하는 최적화 문제를 KKT 조건을 만족하게끔 해결함으로써 primal problem을 dual problem으로 바꿔 풀어도 동일한 최적해를 얻을 수 있는 것이다.

EXAMPLE

지난 논문에서 공부한 C-SVM의 최적화 과정을 KKT 조건을 활용한 라그랑주 승수법을 통해 보임으로써 이해를 돕고자 한다.

C-SVM의 primal problem은 아래와 같다.

$$\min_{w,b,\xi} \quad \frac{1}{2}\|w\|_2^2 + C\sum_{i=1}^{n}\xi_i$$
$$subject\ to \quad \xi_i \geq 0, i = 1, ..., n\ and\ y_i(w^T x_i + b) \geq 1 - \xi_i$$

위 문제를 라그랑주 승수 $\boldsymbol{\alpha}, \boldsymbol{\mu}$를 갖는 dual problem으로 바꾸어보겠다. 이 때 각 라그랑주 승수에 대응하는 조건이 모두 부등식이므로 $\boldsymbol{\alpha}, \boldsymbol{\mu}$는 모두 0보다 커야 한다는 제약조건을 갖는다(KKT).

라그랑주 함수 L은 primal problem을 바탕으로 라그랑주 승수와 제약조건을 곱한 항을 더해 구할 수 있다는 걸 우리는 이제 안다. (convex 꼴이 나와야 하는데 부등식이 greater than의 꼴이므로 목적함수에서 빼는 꼴)

$$L(w, b, \xi_i, \alpha_i, \mu_i) = \frac{1}{2}\|w\|_2^2 + C\sum_{i=1}^{n}\xi_i - \sum_{i=1}^{n}\alpha_i(yi(w^T x_i + b) - 1 + \xi_i) - \sum_{i=1}^{n}\mu_i\xi_i$$

KKT 조건에 따라서 최적화하려는 미지수로 편미분한 식이 0이 된다는 조건을 가지고 $\boldsymbol{\alpha}, \boldsymbol{\mu}$를 구할 수 있다.

$$\frac{\partial L}{\partial w} = 0 \quad \Rightarrow \quad w = \sum_{i=1}^{n}\alpha_i y_i x_i$$

$$\frac{\partial L}{\partial b} = 0 \quad \Rightarrow \quad \sum_{i=1}^{n}\alpha_i y_i = 0$$

$$\frac{\partial L}{\partial \xi_i} = 0 \quad \Rightarrow \quad C - \alpha_i - \mu_i = 0$$

또한 KKT조건의 complementary slackness 조건을 따르려면 아래를 추가로 만족해야 한다.

$$\mu_i \xi_i = 0, \quad i = 1, ..., n$$
$$\alpha_i (y_i (x^T b) - 1 + \xi_i) = 0, \quad i = 1, .., n$$

따라서 w, b, ξ, α, μ가 모두 KKT 조건을 만족하도록 정했기 때문에 dual problem의 최적해인 α, μ를 구하는 과정만으로 primal problem을 푸는 것과 같은 효과를 낼 수 있다.

▼ 참고

KKT 조건

https://ratsgo.github.io/convex%20optimization/2018/01/26/KKT/

Duality

https://ratsgo.github.io/convex%20optimization/2018/01/25/duality/

라그랑주 승수법 (Lagrange Multiplier Method)

라그랑주 승수법 (Lagrange multiplier method)은 프랑스의 수학자 조세프루이 라그랑주 (Joseph-Louis Lagrange)가 제약 조건이 있는 최적화 문제를 풀기 위해 고안한 방법이다. 라그랑주 승수법은 어떠한 문제의 최

https://untitledtblog.tistory.com/96